(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 785 280 A2

(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
23.07.1997 Bulletin 1997/30

(51) Int Cl.^{6}: C12Q 1/68

(21) Application number: 96308616.0

(22) Date of filing: 28.11.1996

(84) Designated Contracting States:
DE FR GB IT NL

(30) Priority: 29.11.1995 US 563762

(71) Applicant: Affymetrix, Inc.
Santa Clara, CA 95051 (US)

(72) Inventors:
- Sapolsky, Ronald
Palo Alto, California (US)
- Lipshutz, Robert J.
Palo Alto, California (US)

(74) Representative: Nash, David Allan
Haseltine Lake & Co.,
Imperial House,
15-19 Kingsway
London WC2B 6UD (GB)

(54) Polymorphism detection

(57) The present invention generally provides a rapid efficient method for analyzing polymorphic or biallelic markers, and arrays for carrying out these analyses. In general, the methods of the present invention employ arrays of oligonucleotide probes that are complementary to target nucleic acids which correspond to the marker sequences of an individual. The probes are typically arranged in detection blocks, each block being capable of discriminating the three genotypes for a given marker, e.g., the heterozygote or either of the two homozygotes. The method allows for rapid, automatable analysis of genetic linkage to even complex polygenic traits.

EP 0 785 280 A2

Description

BACKGROUND OF THE INVENTION

The relationship between structure and function of macromolecules is of fundamental importance in the understanding of biological systems. These relationships are important to understanding, for example, the functions of enzymes, structural proteins and signalling proteins, ways in which cells communicate with each other, as well as mechanisms of cellular control and metabolic feedback.

Genetic information is critical in continuation of life processes. Life is substantially informationally based and its genetic content controls the growth and reproduction of the organism and its complements. The amino acid sequences of polypeptides, which are critical features of all living systems, are encoded by the genetic material of the cell. Further, the properties of these polypeptides, e.g., as enzymes, functional proteins, and structural proteins, are determined by the sequence of amino acids which make them up. As structure and function are integrally related, many biological functions may be explained by elucidating the underlying structural features which provide those functions, and these structures are determined by the underlying genetic information in the form of polynucleotide sequences. Further, in addition to encoding polypeptides, polynucleotide sequences also can be involved in control and regulation of gene expression. It therefore follows that the determination of the make-up of this genetic information has achieved significant scientific importance.

As a specific example, diagnosis and treatment of a variety of disorders may often be accomplished through identification and/or manipulation of the genetic material which encodes for specific disease associated traits. In order to accomplish this, however, one must first identify a correlation between a particular gene and a particular trait. This is generally accomplished by providing a genetic linkage map through which one identifies a set of genetic markers that follow a particular trait. These markers can identify the location of the gene encoding for that trait within the genome, eventually leading to the identification of the gene. Once the gene is identified, methods of treating the disorder that result from that gene, i.e., as a result of overexpression, constitutive expression, mutation, underexpression, etc., can be more easily developed.

One class of genetic markers includes variants in the genetic code termed "polymorphisms." In the course of evolution, the genome of a species can collect a number of variations in individual bases. These single base changes are termed single-base polymorphisms. Polymorphisms may also exist as stretches of repeating sequences that vary as to the length of the repeat from individual to individual. Where these variations are recurring, e.g., exist in a significant percentage of a population, they can be readily used as markers linked to genes involved in mono- and polygenic traits. In the human genome, single-base polymorphisms occur roughly once per 300 bp. Though many of these variant bases appear too infrequently among the allele population for use as genetic markers (i.e., ≤1%), useful polymorphisms (e.g., those occurring in 20 to 50 % of the allele population) can be found approximately once per kilobase. Accordingly, in a human genome of approximately 3 Gb, one would expect to find approximately 3,000,000 of these "useful" polymorphisms.

The use of polymorphisms as genetic linkage markers is thus of critical importance in locating, identifying and characterizing the genes which are responsible for specific traits. In particular, such mapping techniques allow for the identification of genes responsible for a variety of disease or disorder-related traits which may be used in the diagnosis and or eventual treatment of those disorders. Given the size of the human genome, as well as those of other mammals, it would generally be desirable to provide methods of rapidly identifying and screening for polymorphic genetic markers. The present invention meets these and other needs.

SUMMARY OF THE INVENTION

The present invention generally provides arrays and methods useful in screening large numbers of polymorphic markers in a genome. In particular, the present invention provides arrays of oligonucleotide probes for detecting a polymorphism in a target nucleic acid sequence. The array generally comprises at least one detection block of probes which includes first and second groups of probes that are complementary to first and second variants of the target nucleic acid sequence, respectively. The array further comprises third and fourth groups of probes having a sequence identical to the first and second groups, respectively, but including all possible monosubstitutions of positions in the sequence that are within n bases of a base corresponding to the polymorphism, where n is from 0 to 5.

The present invention also provide methods of using the above arrays in screening genomic material for the presence of the polymorphisms.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a schematic illustration of light-directed synthesis of oligonucleotide arrays.

Figure 2A shows a schematic representation of a single oligonucleotide array containing 78 separate detection blocks. Figure 2B shows a schematic illustration of a detection block for a specific polymorphism denoted WI-567. Figure 2B also shows the triplet layout of detection blocks for the polymorphism employing 20-mer oligonucleotide probes having substitutions 7, 10 and 13 bp from the 3' end of the probe. The probes present in the shaded portions of each detection block are shown adjacent to each detection block.

Figure 3 illustrates a tiling strategy for a polymorphism denoted WI-567, and having the sequence 5'-TGCTGCCTT-GGTTC[A/G]AGCCCTCATCTCTTT-3'. A detection block specific for the WI-567 polymorphism is shown with the probe sequences tiled therein listed above. Predicted patterns for both homozygous forms and the heterozygous form are shown at the bottom.

Figure 4 shows a schematic representation of a detection block specific for the polymorphism denoted WI-1959 having the sequence 5'-ACCAAAAATCAGTC[T/C]GGGTAACTGAGAGTG-3' with the polymorphism indicated by the brackets. A fluorescent scan of hybridization of the heterozygous and both homozygous forms are shown in the center, with the predicted hybridization pattern for each being indicated below.

Figure 5 illustrates an example of a computer system used to execute the software of the present invention which determines whether polymorphic markers in DNA are heterozygote, homozygote with a first polymorphic marker or homozygote with a second polymorphic marker.

Figure 6 shows a system block diagram of computer system 1 used to execute the software of the present invention.

Figure 7 shows a probe array including probes with base substitutions at base positions within two base positions of the polymorphic marker. The position of the polymorphic marker is denoted P_0 and which may have one of two polymorphic markers x and y (where x and y are one of A, C, G, or T).

Figure 8 shows a probe array including probes with base substitutions at base positions within two base positions of the polymorphic marker.

Figure 9 shows a high level flowchart of analyzing intensities to determine whether polymorphic markers in DNA are heterozygote, homozygote with a first polymorphic marker or homozygote with a second polymorphic marker.

Figure 10A shows a tiling arrangement of an array tiled for detecting 246 different polymorphic markers, both sense and antisense strands. Each different polymorphism detection block is indicated by a number representing a specific, preidentified polymorphism. Figure 10B shows a fluorescent scan of the array following exposure to fluorescently labelled target sequence.

DETAILED DESCRIPTION OF THE INVENTION

I. General

The present invention generally provides rapid and efficient methods for screening samples of genomic material for polymorphisms, and arrays specifically designed for carrying out these analyses. In particular, the present invention relates to the identification and screening of single base polymorphisms in a sample. In general, the methods of the present invention employ arrays of oligonucleotide probes that are complementary to target nucleic acid sequence segments from an individual (e.g., a human or other mammal) which target sequences include specific identified polymorphisms, or "polymorphic markers." The probes are typically arranged in detection blocks, each block being capable of discriminating the three genotypes for a given marker, e.g., the heterozygote or either of the two homozygotes. The method allows for rapid, automatable analysis of genetic linkage to even complex polygenic traits.

Oligonucleotide arrays typically comprise a plurality of different oligonucleotide probes that are coupled to a surface of a substrate in different known locations. These oligonucleotide arrays, also described as "Genechips™," have been generally described in the art, for example, U.S. Patent No. 5,143,854 and PCT patent publication Nos. WO 90/15070 and 92/10092. These arrays may generally be produced using mechanical synthesis methods or light directed synthesis methods which incorporate a combination of photolithographic methods and solid phase oligonucleotide synthesis methods. See Fodor et al., Science, 251:767-777 (1991), Pirrung et al., U.S. Patent No. 5,143,854 (see *also* PCT Application No. WO 90/15070) and Fodor et al., PCT Publication No. WO 92/10092 and U.S. Patent No. 5,424,186, each of which is hereby incorporated herein by reference. Techniques for the synthesis of these arrays using mechanical synthesis methods are described in, e.g., U.S. Patent No. 5,384,261, incorporated herein by reference in its entirety for all purposes.

The basic strategy for light directed synthesis of oligonucleotides on a VLSIPS™ Array is outlined in Figure 1. The surface of a substrate or solid support, modified with photosensitive protecting groups (X) is illuminated through a photolithographic mask, yielding reactive hydroxyl groups in the illuminated regions. A selected nucleotide, typically in the form of a 3'-O-phosphoramidite-activated deoxynucleoside (protected at the 5' hydroxyl with a photosensitive protecting group), is then presented to the surface and coupling occurs at the sites that were exposed to light. Following capping and oxidation, the substrate is rinsed and the surface is illuminated through a second mask, to expose additional hydroxyl groups for coupling. A second selected nucleotide (e.g., 5'-protected, 3'-O-phosphoramidite-activated

deoxynucleoside) is presented to the surface. The selective deprotection and coupling cycles are repeated until the desired set of products is obtained. Pease et al., *Proc. Natl. Acad. Sci.* (1994) 91:5022-5026. Since photolithography is used, the process can be readily miniaturized to generate high density arrays of oligonucleotide probes. Furthermore, the sequence of the oligonucleotides at each site is known.

II. Identification of Polymorphisms

The methods and arrays of the present invention primarily find use in the identification of so-called "useful" polymorphisms (i.e., those that are present in approximately 20% or more of the allele population). The present invention also relates to the detection or screening of specific variants of previously identified polymorphisms.

A wide variety of methods can be used to identify specific polymorphisms. For example, repeated sequencing of genomic material from large numbers of individuals, although extremely time consuming, can be used to identify such polymorphisms. Alternatively, ligation methods may be used, where a probe having an overhang of defined sequence is ligated to a target nucleotide sequence derived from a number of individuals. Differences in the ability of the probe to ligate to the target can reflect polymorphisms within the sequence. Similarly, restriction patterns generated from treating a target nucleic acid with a prescribed restriction enzyme or set of restriction enzymes can be used to identify polymorphisms. Specifically, a polymorphism may result in the presence of a restriction site in one variant but not in another. This yields a difference in restriction patterns for the two variants, and thereby identifies a polymorphism. Oligonucleotide arrays may also be used to identify polymorphisms. For example, as described in U.S. Patent Application Serial No. 08/485,606, filed June 7, 1995 polymorphisms may be identified using type-IIs endonucleases to capture and amplify ambiguous base sequences adjacent the restriction sites. The captured sequences are then characterized on oligonucleotide arrays. The patterns of these captured sequences are compared from various individuals, the differences being indicative of potential polymorphisms. Alternative array-based methods may also be used to identify polymorphisms, including the methods described in U.S. Patent Application No. 08/629,031, filed April 8, 1996. Briefly, these methods hybridize a target nucleic acid against an appropriately tiled array, e.g., having probes complementary to step-wise segments of the target sequence. The ratio of hybridization intensity of perfectly matched probes to mismatched probes is plotted as a function of the position that is being interogated in the sequence, for each individual screened. Where a polymorphism is present, it yields a discrepency between the data plotted for the individuals, e.g., a point of separation of the two or more individual's plots.

In a preferred aspect, the identification of polymorphisms takes into account the assumption that a useful polymorphism (i.e., one that occurs in 20 to 50% of the allele population) occurs approximately once per lkb in a given genome. In particular, random sequences of a genome, e.g., random 1kb sequences of the human genome such as expressed sequence tags or "ESTs", can be sequenced from a limited number of individuals. When a variant base is detected with sufficient frequency, it is designated a "useful" polymorphism. In practice, the method generally analyzes the same 1 kb sequence from a small number of unrelated individuals, i.e., from 3 to 5 individuals (6 to 10 alleles). where a variant sequence is identified, it is then compared to a separate pool of material from unrelated individuals. Where the variant sequence identified from the first set of individuals is detectable in the pool of the second set, it is assumed to exist at a sufficiently high frequency, e.g., at least about 20% of the allele population, thereby qualifying as a useful marker for genetic linkage analysis.

III. Screening Polymorphisms

Screening polymorphisms in samples of genomic material according to the methods of the present invention, is generally carried out using arrays of oligonucleotide probes. These arrays may generally be "tiled" for a large number of specific polymorphisms. By "tiling" is generally meant the synthesis of a defined set of oligonucleotide probes which is made up of a sequence complementary to the target sequence of interest, as well as preselected variations of that sequence, e.g., substitution of one or more given positions with one or more members of the basis set of monomers, i.e. nucleotides. Tiling strategies are discussed in detail in Published PCT Application No. WO 95/11995, incorporated herein by reference in its entirety for all purposes. By "target sequence" is meant a sequence which has been identified as containing a polymorphism, and more particularly, a single-base polymorphism, also referred to as a "biallelic base. " It will be understood that the term "target sequence" is intended to encompass the various forms present in a particular sample of genomic material, i.e., both alleles in a diploid genome.

In a particular aspect, arrays are tiled for a number of specific, identified polymorphic marker sequences. In particular, the array is tiled to include a number of detection blocks, each detection block being specific for a specific polymorphic marker or set of polymorphic markers. For example, a detection block may be tiled to include a number of probes which span the sequence segment that includes a specific polymorphism. To ensure probes that are complementary to each variant, the probes are synthesized in pairs differing at the biallelic base.

In addition to the probes differing at the biallelic bases, monosubstituted probes are also generally tiled within the

detection block. These monosubstituted probes have bases at and up to a certain number of bases in either direction from the polymorphism, substituted with the remaining nucleotides (selected from A, T, G, C or U). Typically, the probes in a tiled detection block will include substitutions of the sequence positions up to and including those that are 5 bases away from the base that corresponds to the polymorphism. Preferably, bases up to and including those in positions 2 bases from the polymorphism will be substituted. The monosubstituted probes provide internal controls for the tiled array, to distinguish actual hybridization from artifactual cross-hybridization. An example of this preferred substitution pattern is shown in Figure 3.

A variety of tiling configurations may also be employed to ensure optimal discrimination of perfectly hybridizing probes. For example, a detection block may be tiled to provide probes having optimal hybridization intensities with minimal cross-hybridization. For example, where a sequence downstream from a polymorphic base is G-C rich, it could potentially give rise to a higher level of cross-hybridization or "noise," when analyzed. Accordingly, one can tile the detection block to take advantage of more of the upstream sequence. Such alternate tiling configurations are schematically illustrated in Figure 2B, bottom, where the base in the probe that is complementary to the polymorphism is placed at different positions in the sequence of the probe relative to the 3' end of the probe. For ease of discussion, both the base which represents the polymorphism and the complementary base in the probe are referred to herein as the "polymorphic base" or "polymorphic marker."

Optimal tiling configurations may be determined for any particular polymorphism by comparative analysis. For example, triplet or larger detection blocks like those illustrated in Figure 2B may be readily employed to select such optimal tiling strategies.

Arrays may be tiled for one or both strands of the target sequence, i.e., the sequence including the polymorphism. The inclusion of probes that hybridize to both the sense and antisense strands, either on a single array or separate arrays, provides an additional level of verification for a given interogation. Thus, in addition to probes that are complementary to one sdtrand of a target sequence, a detection block will also include probes that are complementary to the antisense strand of the target sequence, and which are therefor complementary to the first group of probes.

Additionally, arrays will generally be tiled to provide for ease of reading and analysis. For example, the probes tiled within a detection block will generally be arranged so that reading across a detection block, the probes are tiled in succession, i.e., progressing along the target sequence one or more bases at a time (See, e.g., Figure 3, middle).

Once an array is appropriately tiled for a given polymorphism or set of polymorphisms, the target nucleic acid is hybridized with the array and scanned. Hybridization and scanning are generally carried out by methods described in, e.g., Published PCT Application Nos. WO 92/10092 and WO 95/11995, and U.S. Patent No. 5,424,186, previously incorporated herein by reference in their entirety for all purposes. In brief, a target nucleic acid sequence which includes one or more previously identified polymorphic markers is amplified by well known amplification techniques, e.g., PCR. Typically, this involves the use of primer sequences that are complementary to the two strands of the target sequence both upstream and downstream from the polymorphism. Asymmetric PCR techniques may also be used, i.e., where an array is tiled for only a sense or antisense strand. Amplified target, generally incorporating a label, is then hybridized with the array under appropriate conditions. Incorporation of a label generally involves incorporating a labeled nucleotide into the amplification reaction, whereby the label is incorporated into the target nucleic acid. Typically useful labels include fluorescent labels coupled to nucleotides, as well as well known binding groups, e.g., biotin, streptavidin and the like, to which a labelled complement may be later bound.

Upon completion of hybridization and washing of the array, the array is scanned to determine the position on the array to which the target sequence hybridizes. The hyridization data obtained from the scan, i.e., in the form of fluorescence intensities or some other detectable label or dye, is then plotted as a function of location on the array.

Although primarily described in terms of a single detection block, e.g., for detection of a single polymorphism, in preferred aspects, the arrays of the invention will include multiple detection blocks, and thus be capable of analyzing multiple, specific polymorphisms. For example, preferred arrays will generally include from about 50 to about 4000 different detection blocks with particularly preferred arrays including from 100 to 3000 different detection blocks.

In alternate arrangements, it will generally be understood that detection blocks may be grouped within a single array or in multiple, separate arrays so that varying, optimal conditions may be used during the hybridization of the target to the array. For example, it may often be desirable to provide for the detection of those polymorphisms that fall within G-C rich stretches of a genomic sequence, separately from those falling in A-T rich segments. This allows for the separate optimization of hybridization conditions for each situation.

IV. Calling

After hybridization and scanning, the hybridization data from the scanned array is then analyzed to identify which variant or variants of the polymorphic marker are present in the sample, or target sequence, as determined from the probes to which the target hybridized, e.g., one of the two homozygote forms or the heterozygote form. This determination is termed "calling" the genotype. Calling the genotype is typically a matter of comparing the hybridization data

for each potential variant, and based upon that comparison, identifying the actual variant (for homozygotes) or variants (for heterozygotes) that are present. In one aspect, this comparison involves taking the ratio of hybridization intensities (corrected for average background levels) for the expected perfectly hybridizing probes for a first variant versus that of the second variant. Where the marker is homozygous for the first variant, this ratio will be a large number, theoretically approaching an infinite value. Where homozygous for the second variant, the ratio will be a very low number, i.e., theoretically approaching zero. Where the marker is heterozygous, the ratio will be approximately 1. These numbers are, as described, theoretical. Typically, the first ratio will be well in excess of 1, i.e., 2, 4, 5 or greater. Similarly, the second ratio will typically be substantially less than 1, i.e., 0.5, 0.2, 0.1 or less. The ratio for heterozygotes will typically be approximately equal to 1, i.e. from 0.7 to 1.5. These ratios can vary based upon the specific sequence surrounding the polymorphism, and can also be adjusted based upon a standard hybridization with a control sample containing the variants of the polymorphism. The ratio may be put on a linear scale bytaking the $\log_{10}$ of the ratio and multiplying the result by 10. This makes it easier to interpret the results of the comparison of the intensities observed.

The quality of a given call for a particular genotype may also be checked. For example, the maximum perfect match intensity can be divided by a measure of the background noise (which may be represented by the standard deviation of the mismatched intensities). Where the ratio exceeds some preselected cut-off point, the call is determined to be good. For example, where the maximum intensity of the expected perfect matches exceeds twice the noise level, it might be termed a good call. In an additional aspect, the present invention provides software for performing the above described comparisons.

Fig. 5 illustrates an example of a computer system used to execute the software of the present invention which determines whether polymorphic markers in DNA are heterozygote, homozygote with a first variant of a polymorphism or homozygote with a second variant of a polymorphism. Fig. 5 shows a computer system 1 which includes a monitor 3, screen 5, cabinet 7, keyboard 9, and mouse 11. Mouse 11 may have one or more buttons such as mouse buttons 13. Cabinet 7 houses a CD-ROM drive 15 or a hard drive (not shown) which may be utilized to store and retrieve software programs incorporating the present invention, digital images for use with the present invention, and the like. Although a CD-ROM 17 is shown as the removable media, other removable tangible media including floppy disks, tape, and flash memory may be utilized. Cabinet 7 also houses familiar computer components (not shown) such as a processor, memory, and the like.

Fig. 6 shows a system block diagram of computer system 1 used to execute the software of the present invention. As in Fig. 5, computer system 1 includes monitor 3 and keyboard 9. Computer system 1 further includes subsystems such as a central processor 102, system memory 104, I/O controller 106, display adapter 108, removable disk 112, fixed disk 116, network interface 118, and speaker 120. Other computer systems suitable for use with the present invention may include additional or fewer subsystems. For example, another computer system could include more than one processor 102 (i.e., a multi-processor system) or a cache memory.

Arrows such as 122 represent the system bus architecture of computer system 1. However, these arrows are illustrative of any interconnection scheme serving to link the subsystems. For example, a local bus could be utilized to connect the central processor to the system memory and display adapter. Computer system 1 shown in Fig. 6 is but an example of a computer system suitable for use with the present invention. Other configurations of subsystems suitable for use with the present invention will be readily apparent to one of ordinary skill in the art.

Fig. 7 shows a probe array including probes with base substitutions at base positions within two base positions of the polymorphic marker. The position of the polymorphic marker is denoted P_0 and which may have one of two variants of the polymorphic markers x and y (where X and y are one of A, C, G, or T). As indicated, at P_{-2} there are two columns of four cells which contain a base substitution two base positions to the left, or 3', from the polymorphic marker. The column denoted by an "x" contains polymorphic marker x and the column denoted by a "y" contains polymorphic marker y.

Similarly, P_{-1} contains probes with base substitutions one base position to the left, or 3', of the polymorphic marker. P_0 contains probes with base substitutions at the polymorphic marker position. Accordingly, the two columns in P_0 are identical. P_1 and P_2 contain base substitutions one and two base positions to the right, or 5', of the polymorphic marker, respectively.

As a hypothetical example, assume a single base polymorphism exists where one allele contains the subsequence TCAAG whereas another allele contains the subsequence TCGAG, where the underlined base indicates the polymorphism in each allele. Fig. 8 shows a probe array including probes with base substitutions at base positions within two base positions of the polymorphic marker. In the first two columns, the cells which contain probes with base A (complementary to T in the alleles) two positions from the left of the polymorphic marker are shaded. They are shaded to indicate that it is expected that these cells would exhibit the highest hybridization to the labeled sample nucleic acid. Similarly, the second two columns have cells shaded which have probes with base G (complementary to C in the alleles) one position to the left of the polymorphic marker.

At the polymorphic marker position (corresponding to P_0 in Fig. 7), there are two columns: one denoted by an "A" and one denoted by a "G". Although, as indicated earlier, the probes in these two columns are identical, the probes

contain base substitutions for the polymorphic marker position. An "N" indicates the cells that have probes which are expected to exhibit a strong hybridization if the allele contains a polymorphic marker A. As will become apparent in the following paragraphs, "N" stands for numerator because the intensity of these cells will be utilized in the numerator of an equation. Thus, the labels were chosen to aid the reader's understanding of the present invention.

A "D" indicates the cells that have probes which are expected to exhibit a strong hybridization if the allele contains a polymorphic marker G. "D" stands for denominator because the intensity of these cells will be utilized in the denominator of an equation. The "n" and "d" labeled cells indicate these cells contain probes with a single base mismatch near the polymorphic marker. As before, the labels indicate where the intensity of these cells will be utilized in a following equation.

Fig. 9 shows a high level flowchart of analyzing intensities to determine whether polymorphic markers in DNA are heterozygote, homozygote with a first polymorphic marker or homozygote with a second polymorphic marker. At step 202, the system receives the fluorescent intensities of the cells on the chip. Although in a preferred embodiment, the hybridization of the probes to the sample are determined from fluorescent intensities, other methods and labels including radioactive labels may be utilized with the present invention. An example of one embodiment of a software program for carrying out this analysis is reprinted in Software Appendix A.

A perfect match (PM) average for a polymorphic marker x is determined by averaging the intensity of the cells at P_0 that have the base substitution equal to x in Fig. 7. Thus, for the example in Fig. 8, the perfect match average for A would add the intensities of the cells denoted by "N" and divide the sum by 2.

A mismatch (MM) average for a polymorphic marker x is determined by averaging the intensity of the cells that contain the polymorphic marker x and a single base mismatch in Fig. 7. Thus, for the example in Fig. 8, the mismatch average for A would be the sum of cells denoted by "n" and dividing the sum by 14.

A perfect match average and mismatch average for polymorphic marker y is determined in a similar manner utilizing the cells denoted by "D" and "d", respectively. Therefore, the perfect match averages are an average intensity of cells containing probes that are perfectly complementary to an allele. The mismatch averages are an average of intensity of cells containing probes that have a single base mismatch near the polymorphic marker in an allele.

At step 204, the system calculates a Ratio of the perfect match and mismatch averages for x to the perfect match and mismatch averages for y. The numerator of the Ratio includes the mismatch average for x subtracted from the perfect mismatch for x. In a preferred embodiment, if the resulting numerator is less than 0, the numerator is set equal to 0.

The denominator of the Ratio includes the mismatch average for y subtracted from the perfect mismatch for y. In a preferred embodiment, if the resulting denominator is less than or equal to 0, the denominator is set equal to a minimum value, i.e., 0.00001.

Once the system has calculated the Ratio, the system calculates DB at step 206. DB is calculated by the equation $DB = 10 * \log_{10} Ratio$. The logarithmic function puts the ratio on a linear scale and makes it easier to interpret the results of the comparison of intensities.

At step 208, the system performs a statistical check on the data or hybridization intensities. The statistical check is performed to determine if the data will likely produce good results. In a preferred embodiment, the statistical check involves testing whether the maximum of the perfect match averages for x or y is at least twice as great as the standard deviation of the intensities of all the cells containing a single base mismatch (i.e., denoted by a "n" or "d" in Fig. 8). If the perfect match average is at least two times greater than this standard deviation, the data is likely to produce good results and this is communicated to the user.

The system analyzes DB at step 210 to determine if DB is approaching $-\infty$, near 0, or approaching $+\infty$. In practice, the DB will typically ot go beyond 50 or -50. If DB is approaching a negative infinity (e.g., -50), the system determines that the sample DNA contains a homozygote with a first polymorphic marker corresponding to x at step 212. If DB is near 0, the system determines that the sample DNA contains a heterozygote corresponding to both polymorphic markers x and y at step 214. Although described as approaching ∞, etc., as described previously, these numbers will generally vary, but are nonetheless indicative of the calls described. If DB is approaching a positive infinity (e.g., +50), the system determines that the sample DNA contains a homozygote with a second polymorphic marker corresponding to y at step 216.

A visual inspection of the Ratio equation in step 204 shows that the numerator should be higher than the denominator if the DNA sample only has the polymorphic marker corresponding to x. Similarly, the denominator should be higher than the numerator if the DNA sample only has a polymorphic marker corresponding to y. If the DNA sample has both polymorphic markers, indicating a heterozygote, the Ratio should be approximately equal to 1 which results in a 0 when the logarithm of the Ratio is calculated.

The equations discussed above illustrate just one embodiment of the present invention. These equations have correctly identified polymorphic markers when a visual inspection would seem to indicate a different result. This may be the case because the equations take into account the mismatch intensities in order to determine the presence or absence of the polymorphic markers. Additional methods may also be employed to properly compare the hybridization

intensities, including, e.g., principal component analysis and the like.

Those of skill in the art, upon reading the instant disclosure will appreciate that a variety of modifications to the above described methods and arrays may be made without deparing from the spirit or scope of the invention. For example, one may select the strand of the target sequence to optimize the ability to call a particular genome. Alternatively, one may analyze both strands, in parallel, to provide greater amounts of data from which a call can be made. Additionally, the analyses, i.e., amplification and scanning may be performed using DNA, RNA, mixed polymers, and the like.

The present invention is further illustrated by the following examples. These examples are merely to illustrate aspects of the present invention and are not intended as limitations of this invention.

V. Examples

Example 1- Chip Tiling

A DNA chip is prepared which contains three detection blocks for each of 78 identified single base polymorphisms or biallelic markers, in a segment of human DNA (the "target" nucleic acid). Each detection block contains probes wherein the identified polymorphism occurs at the position in the target nucleic acid complementary to the 7th, 10th and 13th positions from the 3' end of 20-mer oligonucleotide probes. A schematic representation of a single oligonucleotide array containing all 78 detection blocks is shown in Figure 2A.

The tiling strategy for each block substitutes bases in the positions at, and up to two bases, in either direction from the polymorphism. In addition to the substituted positions, the oligonucleotides are synthesized in pairs differing at the biallelic base. Thus, the layout of the detection block (containing 40 different oligonucleotide probes) allows for controlled comparison of the sequences involved, as well as simple readout without need for complicated instrumentation. A schematic illustration of this tiling strategy within a single detection block is shown in Figure 3, for a specific polymorphic marker denoted WI-567.

Example 2- Detection of Polymorphisms

A target nucleic acid is generated from PCR products amplified by primers flanking the markers. These amplicons can be produced singly or in multiplexed reactions. Target can be produced as ss-DNA by asymmetric PCR from one primer flanking the polymorphism, as ds-DNA, or as RNA transcribed in vitro from promoters linked to the primers. Fluorescent or biotin label is introduced into target directly as dye or biotin-bearing nucleotides. Biotin labelled target is then bound after amplification using dye-streptavidin complexes to incorporated biotin containing nucleotides. In DNA produced by symetric or asymetric PCR fluorescent dye is linked directly to the 5' end of the primer.

Hybridization of target to the arrays tiled in Example 1, and subsequent washing are carried out with standard solutions of salt (SSPE, TMAC1) and nonionic detergent (Triton-X100), with or without added organic solvent (formamide). Targets and markers generating strong signals are washed under stringent hybridization conditions (37-40°C; 10% formamide; 0.25xSSPE washes) to give highly discriminating detection of the genotype. Markers giving lower hybridization intensity are washed under less stringent conditions (≤30°C; 3M TMAC1, or 6xSSPE; 6x and 1x SSPE washes) to yield highly discriminating detection of the genotype.

Detection of one polymorphic marker is illustrated in Figure 3. Specifically, a typical detection block is shown for the polymorphism denoted WI-1959, having the sequence 5'-ACCAAAAATCAGTC[T/C]GGGTAACTGAGAGTG-3' with the polymorphism indicated by the brackets (Figure 3, top), for which all three genotypes are available (T/C heterozygote, C/C homozygote and T/T homozygote). The expected hybridization pattern for the homozygote and heterozygote targets are shown in Figure 3, bottom. Three chips were tiled with each chip including the illustrated detection block. Each block contained probes having the substituted bases at the 7th, 10th and 13th positions from the 3' end of 20-mer oligonucleotide probes (20/7, 20/10 and 20/13, respectively). These alternate detection blocks were tiled to provide a variety of sequences flanking the polymorphism itself, to ensure at least one detection block hybridizing with a sufficiently low background intensity for adequate detection.

Fluorouracil containing RNA was synthesized from a T7 promoter on the upstream primer, hybridized to the detection array in 6xSSPE + Triton-X100 at 30°C, and washed in 0.25xSSPE at room temperature. As shown in the scan Figure 3, middle, fluorescent scans of the arrays correctly identified the 5 homozygote or 10 heterozygote features.

While the foregoing invention has been described in some detail for purposes of clarity and understanding, it will be clear to one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the invention. All publications and patent documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication or patent document were so individually denoted.

APPENDIX A

SOFTWARE APPENDIX

```
# fullcel.awk                                                #
# takes input from a POLYchip CEL file (115 x 130) and       #
# extracts ratio information for every block on the chip #

BEGIN{
 ratpatcutoff = 1.2
 pattoggle = "yes"
 base[0]="T"
 base[1]="G"
 base[2]="C"
 base[3]="A"
 name[0,0] = "WI-563"
 hex[0,0] = "CTAGCC"
 name[1,0] = "WI-567"
 hex[1,0] = "TCAGAG"
 name[2,0] = "WI-597"
 hex[2,0] = "TGGATA"
 name[3,0] = "WI-681"
 hex[3,0] = "AACTAA"
 name[4,0] = "WI-801"
 hex[4,0] = "CTTGAG"
 name[5,0] = "WI-802"
 hex[5,0] = "CATCCT"
 name[6,0] = "WI-1099"
 hex[6,0] = "CAGATA"
 name[7,0] = "WI-1147"
 hex[7,0] = "ACGAGC"
 name[8,0] = "WI-1325"
 hex[8,0] = "CTCTAC"
 name[9,0] = "WI-1417"
 hex[9,0] = "GTCTTT"
 name[0,1] = "WI-1796"
 hex[0,1] = "AAAGTG"
 name[1,1] = "WI-1825"
 hex[1,1] = "GTCTTC"
 name[2,1] = "WI-1879"
 hex[2,1] = "TACTGT"
 name[3,1] = "WI-1888"
 hex[3,1] = "ATGACA"
 name[4,1] = "WI-1912"
 hex[4,1] = "TTCTTT"
 name[5,1] = "WI-1959"
 hex[5,1] = "TCTCGG"
 name[6,1] = "WI-1741"
 hex[6,1] = "GAAGGC"
 name[7,1] = "WI-1760"
 hex[7,1] = "ACCACA"
 name[8,1] = "WI-1799"
 hex[8,1] = "TCGATA"
 name[9,1] = "WI-1973"
 hex[9,1] = "CAAGAG"
 name[0,2] = "WI-1980"
 hex[0,2] = "AACTGA"
 name[1,2] = "WI-2015"
```

```
hex[1,2] = "GACTGT"
name[2,2] = "WI-2664"
hex[2,2] = "GGAGAG"
name[3,2] = "WI-4013"
hex[3,2] = "CTAGTG"
name[4,2] = "WI-7567"
hex[4,2] = "TAGTGA"
name[5,2] = "WI-11595"
hex[5,2] = "TAGAGC"
name[6,2] = "CM4.16"
hex[6,2] = "GATAAT"
name[7,2] = "WI-6704"
hex[7,2] = "ACTCCA"
name[8,2] = "WI-6731"
hex[8,2] = "GGCACA"
name[9,2] = "WI-6787"
hex[9,2] = "ACAGTT"
name[0,3] = "WI-6910"
hex[0,3] = "TAGTTG"
name[1,3] = "WI-9518"
hex[1,3] = "TTGATT"
name[2,3] = "ADH3"
hex[2,3] = "ATAGTT"
name[3,3] = "AGT"
hex[3,3] = "GACTGG"
name[4,3] = "ALDOB-1"
hex[4,3] = "TTCTGG"
name[5,3] = "ALDOB-2"
hex[5,3] = "CCAGAT"
name[6,3] = "APOB"
hex[6,3] = "ACTCCT"
name[7,3] = "APOE(152T/C)"
hex[7,3] = "TGTCGC"
name[8,3] = "APOE(290T/C)"
hex[8,3] = "AGTCGC"
name[9,3] = "ARSB"
hex[9,3] = "TCGATG"
name[0,4] = "AT1a"
hex[0,4] = "CTTCCC"
name[1,4] = "AT1b"
hex[1,4] = "GCACTT"
name[2,4] = "BCL2"
hex[2,4] = "ACGAGG"
name[3,4] = "BRCA1a"
hex[3,4] = "CATCTG"
name[4,4] = "BRCA1b"
hex[4,4] = "AGAGAG"
name[5,4] = "BRCA1c"
hex[5,4] = "GAAGAG"
name[6,4] = "D3S2"
hex[6,4] = "CCAGGT"
name[7,4] = "D3S11"
hex[7,4] = "TCTGRR"
name[8,4] = "D3S12"
hex[8,4] = "CCAGGG"
name[9,4] = "DRD2"
hex[9,4] = "CACTGG"
name[0,5] = "FABP2"
hex[0,5] = "GCGACT"
```

```
name[1,5] = "GCK"
hex[1,5] = "GAGACA"
name[2,5] = "HT2"
hex[2,5] = "CTGTGG"
name[3,5] = "HT4"
hex[3,5] = "TGCAAT"
name[4,5] = "HT5"
hex[4,5] = "ACTCGA"
name[5,5] = "IGF2"
hex[5,5] = "GGGACC"
name[6,5] = "IGHV4-6"
hex[6,5] = "TCTCGA"
name[7,5] = "INS"
hex[7,5] = "TCTACC"
name[8,5] = "LDLR"
hex[8,5] = "GGCTAA"
name[9,5] = "LF79"
hex[9,5] = "CCAGGG"
name[0,6] = "LPL"
hex[0,6] = "AGCTAG"
name[1,6] = "MCC"
hex[1,6] = "GCCTGA"
name[2,6] = "METH"
hex[2,6] = "CCCTGG"
name[3,6] = "NRAMP"
hex[3,6] = "CAGATG"
name[4,6] = "PAR"
hex[4,6] = "ACATTG"
name[5,6] = "Per/RDS"
hex[5,6] = "GAAGGA"
name[6,6] = "PPP3R1"
hex[6,6] = "GACTAA"
name[7,6] = "RDS"
hex[7,6] = "AGGACG"
name[8,6] = "s14544"
hex[8,6] = "TCTGCT"
name[9,6] = "S180A"
hex[9,6] = "GGCATG"
name[0,7] = "TcR-CA1"
hex[0,7] = "TGCGGT"
name[1,7] = "TcR-CB22"
hex[1,7] = "GGCTGG"
name[2,7] = "TcR-CB23"
hex[2,7] = "CTCTAG"
name[3,7] = "TcR-CB24"
hex[3,7] = "GTGATG"
name[4,7] = "TcR-CB25"
hex[4,7] = "GTAGCC"
name[5,7] = "TcR-CB27"
hex[5,7] = "ACCTTA"
name[6,7] = "VB12a"
hex[6,7] = "ACAGTG"
name[7,7] = "VB12b"
hex[7,7] = "CACTCA"
bkgsum = 0
bkgnum = 0
}
```

```
  {
   readthis = 1
   if ($1 ~ /[A-Za-z]/ || $2 ~ /[A-Za-z]/) readthis = 0
   if (readthis == 1) rawdata[$1,$2] = $3
   if ($1>2 && $2>4) if ($1<112 && $2<124) if ($1<90 || $2<109)
          {
           px = int(($1-3)/11)
           py = int(($2-5)/15)
           pxo = (11*px)+3
           pyo = (15*py)+5
           mx = $1-pxo
           by = $2-pyo
           block = 3*(int(by/5))+7
           if (by%5 != 4 && mx != 10)
                  {
                   sb = base[by%5]
                   sig[px,py,block,sb,mx] = $3
                  }
          if (by%5 == 4 || mx == 10)
                  {
                   bkgsum += $3
                   bkgnum++
                  }
          }
}
END{
 printf ("background = %5.2f\n", bkgsum/bkgnum)
 printf "MARKER\tBSTBLK\tRATIO\t\tDB\tCHECK\t\tPATRAT\n"
 for (py=0;py<8;py++) for (px=0;px<10;px++) if (py < 7 || px < 8)
    {
     m[0] = substr(hex[px,py],1,1)
     m[1] = substr(hex[px,py],1,1)
     m[2] = substr(hex[px,py],2,1)
     m[3] = substr(hex[px,py],2,1)
     m[4] = substr(hex[px,py],3,2)
     m[5] = substr(hex[px,py],3,2)
     m[6] = substr(hex[px,py],5,1)
     m[7] = substr(hex[px,py],5,1)
     m[8] = substr(hex[px,py],6,1)
     m[9] = substr(hex[px,py],6,1)
     center = substr(hex[px,py],3,1)"/"substr(hex[px,py],4,1)
     pentamer = m[0]" "m[2]"["center"]"m[6]" "m[8]
     header = "["px+1","py+1"] " name[px,py] "\n" pentamer "\n"
     headprint = 0
        {
         for (j=0;j<=2;j++)
                {
                 block = (3*j)+7
                 num2 = 0
                 den2 = 0
                 num1 = 0
                 den1 = 0
                 x2 = 0
                 n1 = 0
                 n2 = 0
                 for (f=0;f<5;f++)
                        {
                         maxhi[px,py,block,f] = 0
                         for (g=0;g<4;g++) maxlo[px,py,block,g,f] = 0
                        }
```

```
for (k=0;k<=9;k++) for (b=0;b<=3;b++)
       {
        z = int(k/2)
        signal = sig[px,py,block,base[b],k]
        omit = 0
        if (m[k] - base[b]) omit = 1
        if (omit == 1)
               {
                q = maxhi[px,py,block,z]
                if (signal > q) maxhi[px,py,block,z]=signal
               }
        if (omit == 0)
               {
                q = maxlo[px,py,block,b,z]
                if (signal > q) maxlo[px,py,block,b,z]=signal
                if (k%2 == 0)
                       {
                        num2 += signal
                        x2 += (signal)^2
                        n1++
                       }
                if (k%2 == 1)
                       {
                        den2 += signal
                        x2 += (signal)^2
                        n2++
                       }
               }
        if (omit == 1) if (k==4 || k==5)
               {
                 if (base[b] == substr(hex[px,py],3,1))
                        {
                         num1 += signal
                        }
                 if (base[b] == substr(hex[px,py],4,1))
                        {
                         den1 += signal
                        }
               }
       }
maxhisum = 0
for (f=0;f<5;f++)
       {
        maxhisum += maxhi[px,py,block,f]
       }
maxhiav = maxhisum/5
maxlosum = 0
for (g=0;g<5;g++) for (v=0;v<4;v++)
       {
        maxlosum += maxlo[px,py,block,v,g]
       }
maxloav = maxlosum/14
maxrat = maxhiav/maxloav
num = ((num1/2)-(num2/n1))
if (num < 0) num = 0
den = ((den1/2)-(den2/n2))
if (den <= 0) den = 0.001
ratio = num/den
max = num1/2
```

```
                        if (den1/2 > max) max = den1/2
                        n = n1+n2
                        stdvxnum = ((n*x2)-(num2+den2)^2)
                        if (stdvxnum < 0) stdvx = 0
                        stdvx = (stdvxnum/(n^2))^(0.5)
                        if (maxrat > ratpatcutoff || pattoggle == 'no')
                                {
                                 if (headprint == 0)
                                         {
                                          printf header
                                          headprint = 1
                                         }
                                 printf '\t20/'block'\t'
                                 printf ('%1.2f\t', ratio)
                                 if (ratio < 10000) printf '\t'
                                 rat = ratio
                                 if (ratio == 0) rat = .00001
                                 lograt = log(rat)/log(10)
                                 printf ('%2.2f\t', 10*lograt)
                                 printf ('%2.2f', max/stdvx)
                                 if (max/stdvx < 2) printf '\tFAIL\t'
                                 if (max/stdvx >= 2) printf '\t\t'
                                 printf ('%2.2f', maxrat)
                                 if (maxrat > ratpatcutoff) printf '\t*GOODPAT*'
                                 printf '\n'
                                }
                       }
               }
         }
    }
```

Claims

1. An array of oligonucleotide probes for detecting a polymorphism in a target nucleic acid sequence, said array comprising at least one detection block of probes said detection block including first and second groups of probes that are complementary to said target nucleic acid sequence having first and second variants of said polymorphism, respectively, and further comprising third and fourth groups of probes, said third and fourth groups of probes having a sequence identical to said first and second groups of probes, respectively, except that said third and fourth groups of probes include all possible monosubstitutions of positions in said sequence that are within n bases of a base in said sequence that is complementary to said polymorphism, wherein n is from 0 to 5.

2. The array of claim 1, wherein n is 2.

3. The array of claim 1, wherein said first and second groups of probes comprise a plurality of different probes that are complementary to overlapping portions of said target nucleic acid sequence.

4. The array of claim 1, wherein said detection block further comprises said third and fourth groups of probes wherein said monosubstitutions occur at a plurality of distances from a 3' end of said probes.

5. The array of claim 1, wherein said detection block includes between about 8 and 88 different probes.

6. The array of claim 1, comprising between 1 and 1,000 different detection blocks, each of said detection blocks including probes complementary to first and second variants of a different polymorphism in said target nucleic acid sequence.

7. The array of claim 1, wherein each of said detection blocks further comprises fifth and sixth groups of probes, said fifth and sixth groups of probes being complementary to said first and second groups of probes, respectively.

8. A method of identifying whether a target nucleic acid sequence includes a polymorphic variant, comprising:

   hybridizing said target nucleic acid sequence to an array of oligonucleotide probes, said array comprising at least one detection block of probes said detection block including first and second groups of probes that are complementary to said target nucleic acid sequence having first and second variants of said polymorphism, respectively, and further comprising third and fourth groups of probes, said third and fourth groups of probes having a sequence identical to said first and second groups of probes, respectively, except that said third and fourth groups of probes include all possible monosubstitutions of positions in said sequence that are within n bases of a base in said sequence that is complementary to said polymorphism, wherein n is from 0 to 5; and determining which of said first and second groups of probes hybridizes with said target nucleic acid sequence to identify said polymorphic variant.

9. The method of claim 8, wherein said target nucleic acid comprises a detectable label.

10. The method of claim 9, wherein said detectable label is a fluorescent group.

11. The method of claim 9, wherein said label is a binding group.

12. The method of claim 11, wherein said binding group is selected from biotin, avidin and streptavidin.

13. The method of claim 8, wherein said detection block includes fifth and sixth groups of probes, said fifth and sixth groups of probes being complementary to first and second variants of an antisense strand of said target sequence.

14. The method of claim 8, wherein said step of determining comprises:

   calculating a ratio of hybridization intensity of said target nucleic acid to said first group of probes versus hybridization intensity of said target nucleic acid to said second group of probes;
   and identifying a homozygote for said first variant when said ratio is greater than 2, a homozygote for said second variant when said ratio is less than 0.5, and a heterozygote when said ratio is between about 0.7 and 1.5.

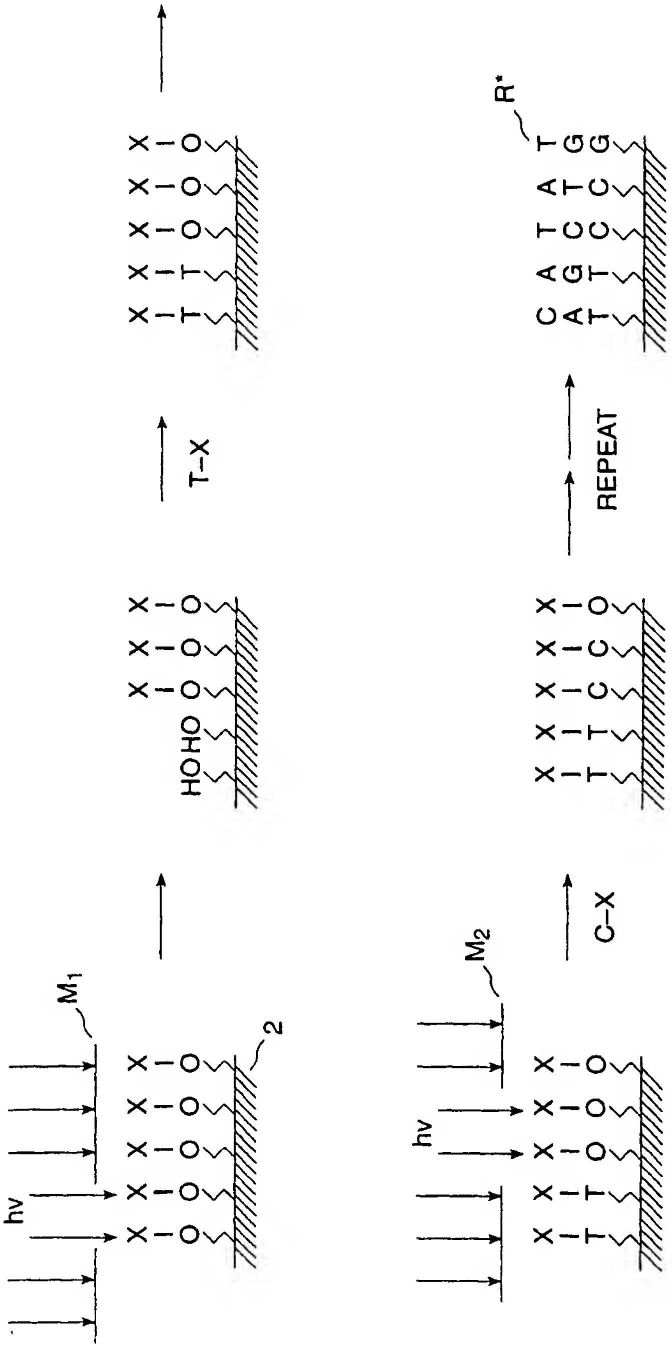


FIG. 1

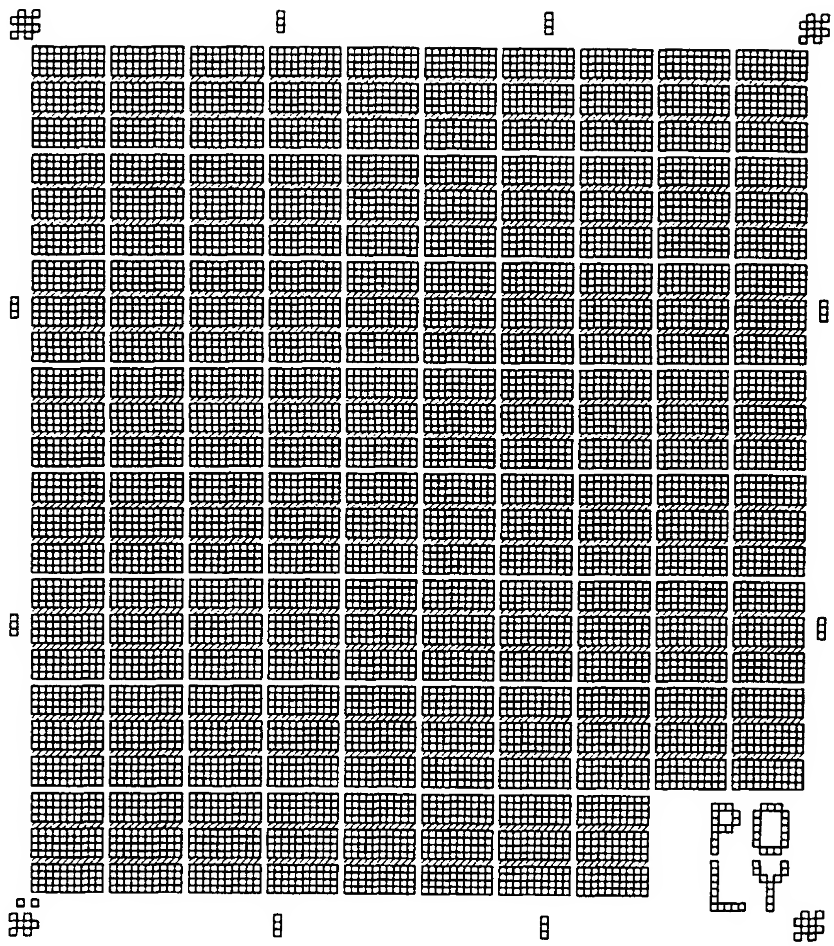


FIG. 2A

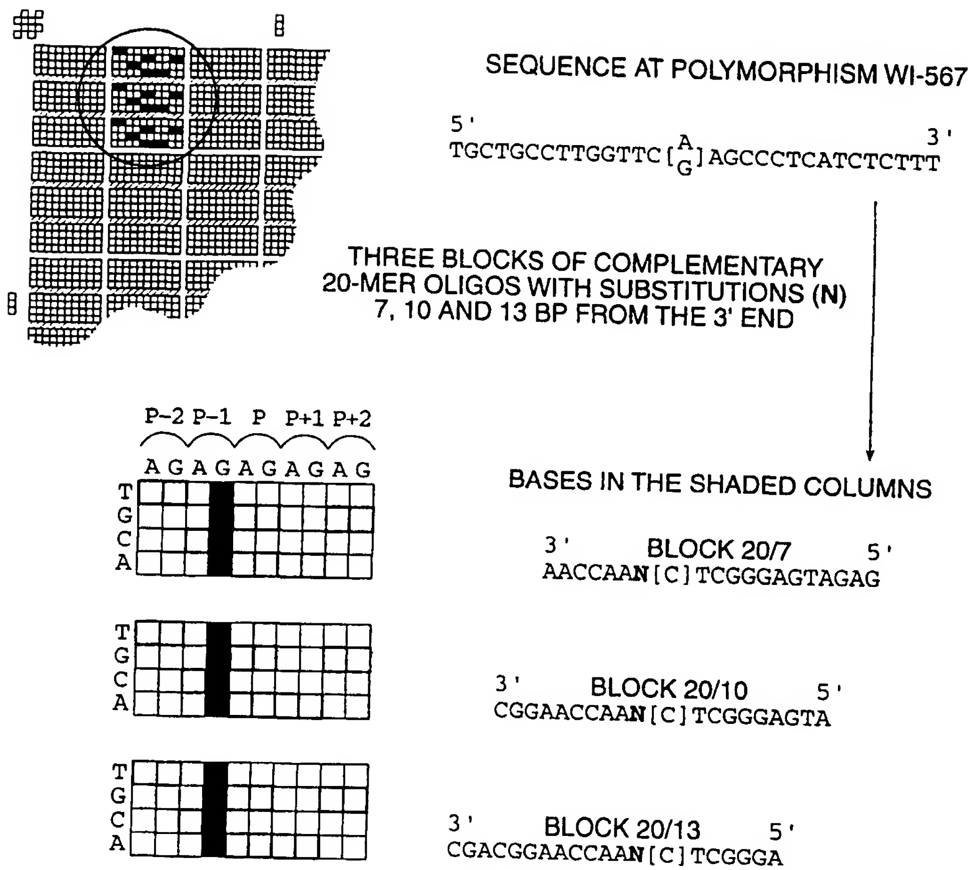
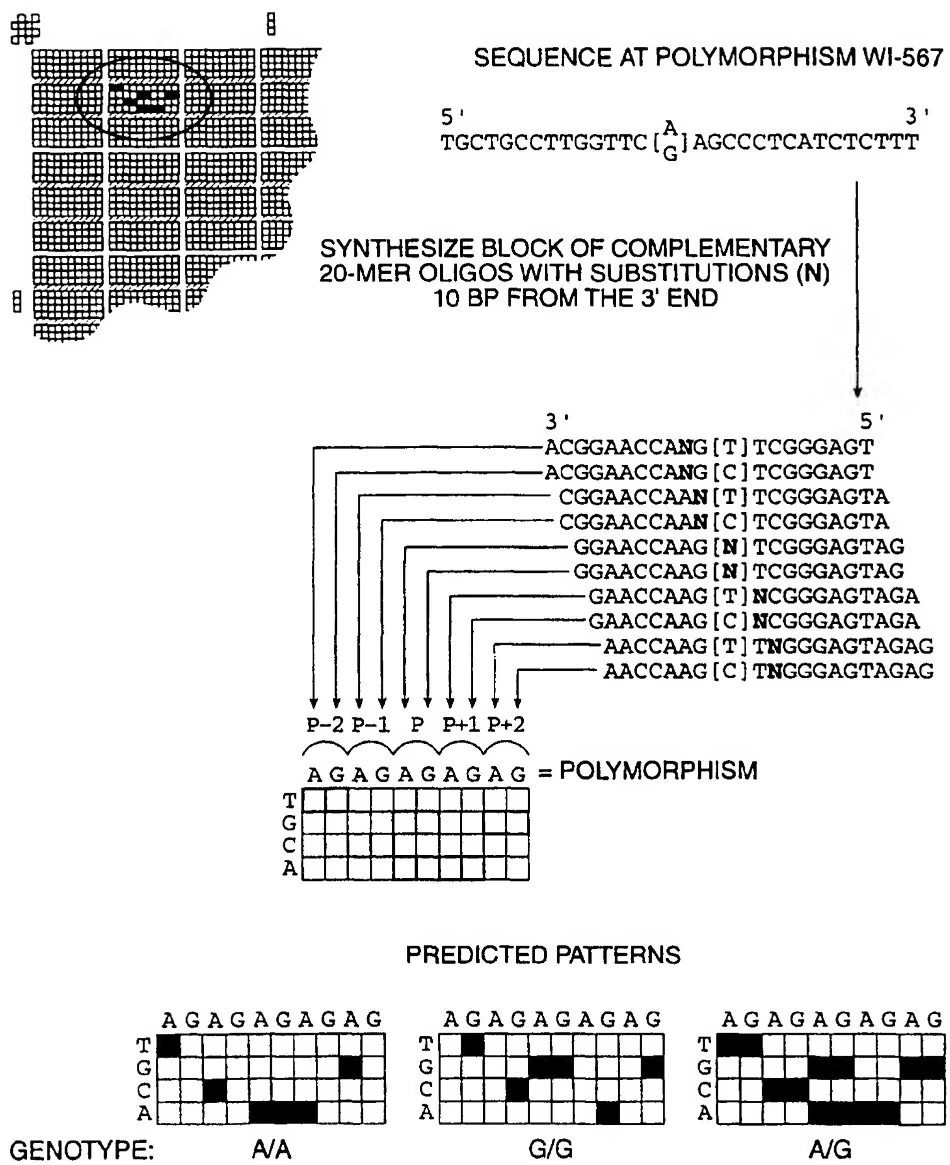
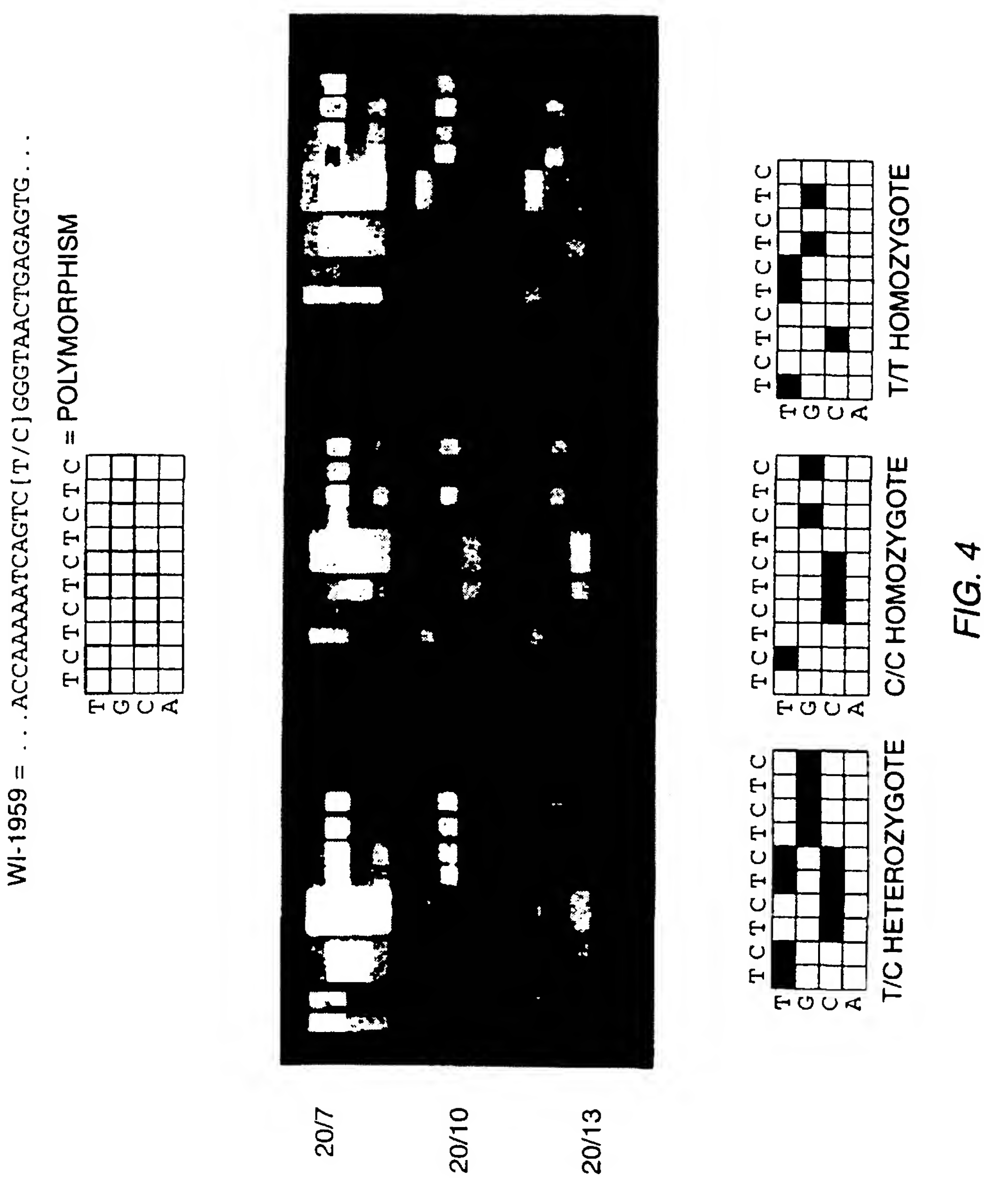

SEQUENCE AT POLYMORPHISM WI-567
5'
3'
TGCTGCCTTGGTTC [A/G] AGCCCTCATCTCTTT
THREE BLOCKS OF COMPLEMENTARY
20-MER OLIGOS WITH SUBSTITUTIONS (N)
7, 10 AND 13 BP FROM THE 3' END
P-2 P-1 P P+1 P+2
A G A G A G A G A G
T
G
C
A
BASES IN THE SHADED COLUMNS
3' BLOCK 20/7 5'
AACCAAN [C] TCGGGAGTAGAG
3' BLOCK 20/10 5'
CGGAACCAAN [C] TCGGGAGTA
3' BLOCK 20/13 5'
CGACGGAACCAAN [C] TCGGGA

FIG. 2B



SEQUENCE AT POLYMORPHISM WI-567
5'
3'
TGCTGCCTTGGTTC [A/G] AGCCCTCATCTCTTT
SYNTHESIZE BLOCK OF COMPLEMENTARY
20-MER OLIGOS WITH SUBSTITUTIONS (N)
10 BP FROM THE 3' END
3'
5'
ACGGAACCANG [T] TCGGGAGT
ACGGAACCANG [C] TCGGGAGT
CGGAACCAAN [T] TCGGGAGTA
CGGAACCAAN [C] TCGGGAGTA
GGAACCAAG [N] TCGGGAGTAG
GGAACCAAG [N] TCGGGAGTAG
GAACCAAG [T] NCGGGAGTAGA
GAACCAAG [C] NCGGGAGTAGA
AACCAAG [T] TNGGGAGTAGAG
AACCAAG [C] TNGGGAGTAGAG
P-2 P-1 P P+1 P+2
A G A G A G A G A G = POLYMORPHISM
T
G
C
A
PREDICTED PATTERNS
A G A G A G A G A G
A G A G A G A G A G
A G A G A G A G A G
T
G
C
A
GENOTYPE:
A/A
G/G
A/G

FIG. 3



WI-1959 = . . . ACCAAAAATCAGTC [T / C] GGGTAACTGAGAGTG . . .
T C T C T C T C T C T C = POLYMORPHISM
T
G
C
A
20/7
20/10
20/13
T C T C T C T C T C T C
T/C HETEROZYGOTE
C/C HOMOZYGOTE
T/T HOMOZYGOTE

FIG. 4

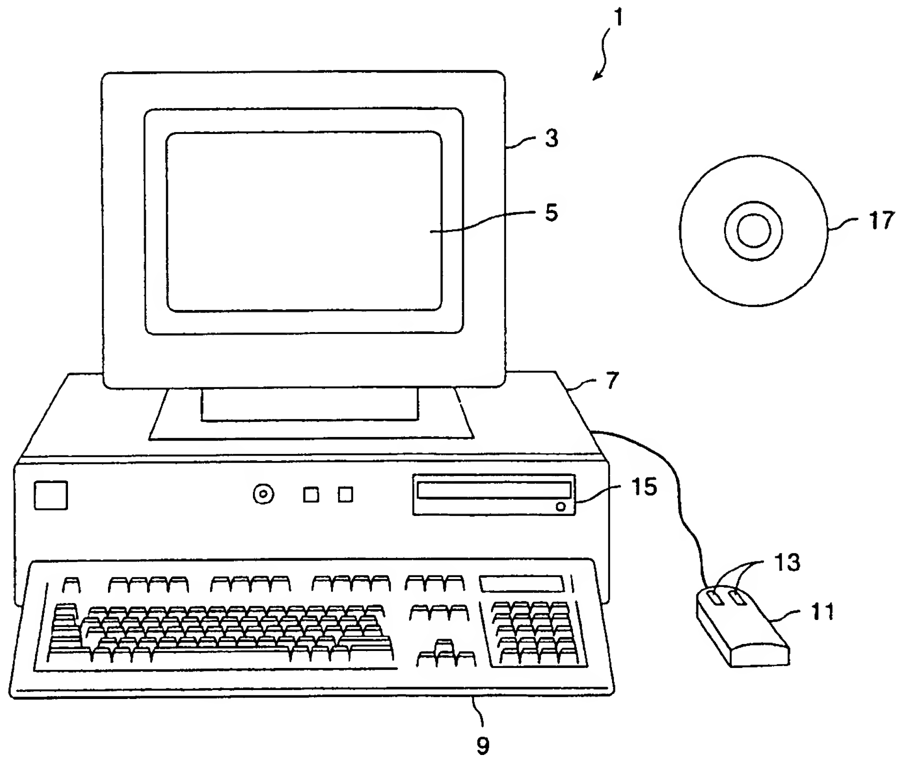

1
3
5
17
7
15
13
11
9

FIG. 5

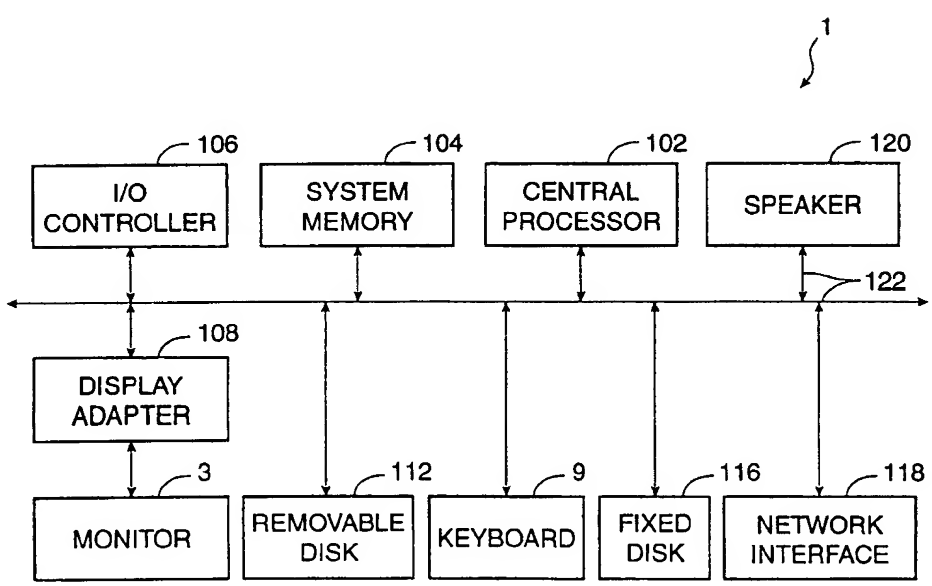

1
106
I/O CONTROLLER
104
SYSTEM MEMORY
102
CENTRAL PROCESSOR
120
SPEAKER
122
108
DISPLAY ADAPTER
3
MONITOR
112
REMOVABLE DISK
9
KEYBOARD
116
FIXED DISK
118
NETWORK INTERFACE

FIG. 6

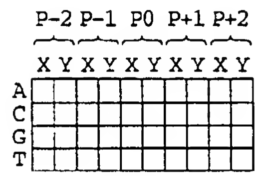


FIG. 7

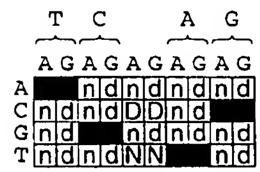


FIG. 8

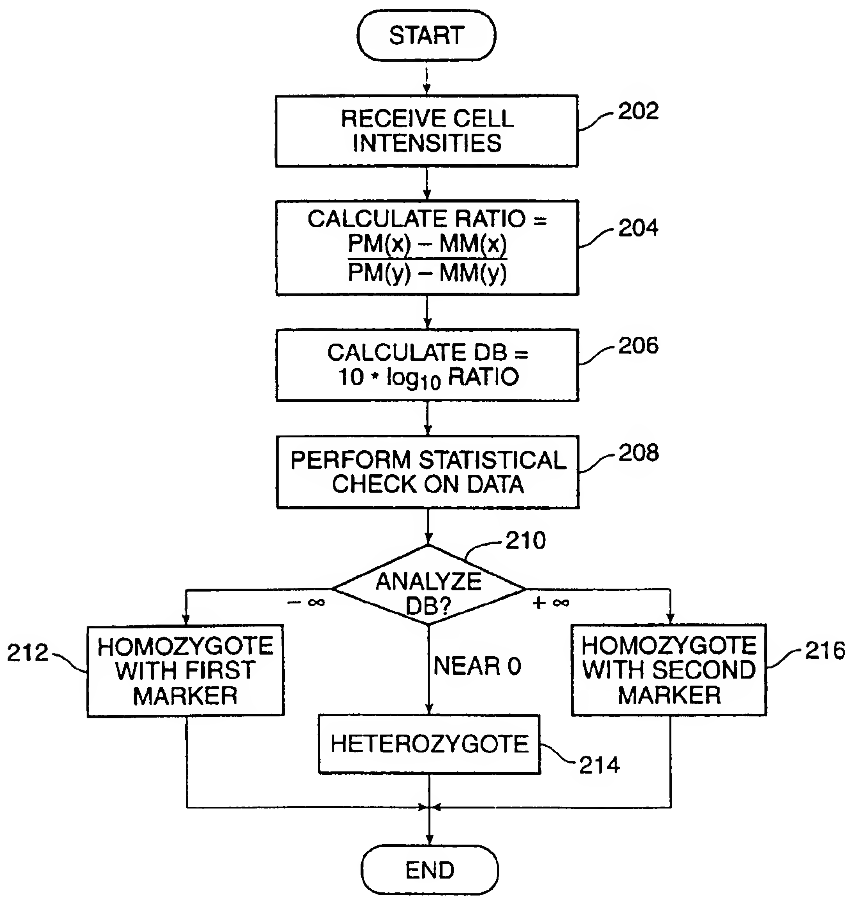


FIG. 9

	From:		3	23	43	63	83	103	123	143	163	183	203	223	243
	To:		22	42	62	82	102	122	142	162	182	202	222	242	262
From:	To:		1	2	3	4	5	6	7	8	9	10	11	12	13
5	17	1	1099	1147	11595	1306	**1325**	1341	1403	1417	1731	1741	1760	178	1796
18	30	2	1799	1825	1879	1888	1912	1959	1973	198	1980	**2015**	2026	205	2532
31	43	3	2664	276	4013	563	567	597	6087	681	6996	7021	7056	7091m1	7115
44	56	4	7146	7169	7172	**7178**	7182	7191	7199	7216	7220	7226iA	7233	7249.1.	7281iA
57	69	5	7282	**7301m1**	7301m5	7314	7321	**7331**	7336	7338m1	7338m2	**7372**	7384	754	7388
70	82	6	7395	7449	7534	7543	7555	7567	7569	7574	7576	7577m6	**7619m4**	**7620m6**	7626m3
83	95	7	7686	7689	7690	7703	7743m1	**7743m2**	7765	7773m1	7774	7785iT	7789	7790	**7795m1**
96	108	8	7814	7830m1	7865m1	7865m2	7867	7868	7870	7889iG	7894	7900	7901m1	7901m2	7926m2
109	121	9	7936	7974	7963	7972	7992	8004	801	802	8021m1	8024	8118m3	8171a	8171b
122	134	10	8288	8314	8332	8371	8378	8382	8411	8458	8461m1	**8461m2**	8505	998	**AGT**
135	147	11	ACP1	ADH3	AK-168	**ALB**	ALDOBm1	ALPP	**APOA2**	APOB	**APOE1**	APOE2	ARSB	AT1a	AT1b
148	160	12	BA51	BCL2	BCR	BCRA1a	BCRA1b	BCRA1c	BTK	C1R	C6	**CA1**	CA2	CB22	CB23
161	173	13	CB24	CB25	CB27	CM4.16	F9	CYP2D6	D11S1873	D17S33M1	D17S33m2	D18S8	D3S12	D3S2	D3S11dTT
174	186	14	D7S399	**DMm1**	DMm2	DRD1	DRD2	DRD3	ERBB2	ETS2m1	FABP2	**GCDH**	GCK	**GH1**	GNAT2
187	199	15	WI-6704	WI-6731	WI-6787	WI-6910	WI-9518	HSD3B1	HT2	HT4	HT5	IGF2	IGFBP1	IGHV	IL1A
200	212	16	IL1B	INS	IPW	ITGB2	ITIH1	KRAS2	KRT10m1	KRT10m2	KRT8m1	LDLR	LF79	LIP	LMP2
213	225	17	LPL	**MCC**	METH	NF1	NFKBI	NPPA	NRAS	PAI1dG	PAR	**PerRDS**	PPP3R1	PTH	PXMP1
226	238	18	RDS	s14544	S180A	SAG	SNRPN	SPTB	THRB	TYR	**TYRP1**	VB12m1	VB12m2	VTN	WI-1147
239	251	19	WI-1305	**WI-1327**	WI-1349	WI-1732	WI-1803	WI-1943	WI-1960	WI-2032	WI-2054	WI-562	WI-867	**WI-945**	

FIG. 10A

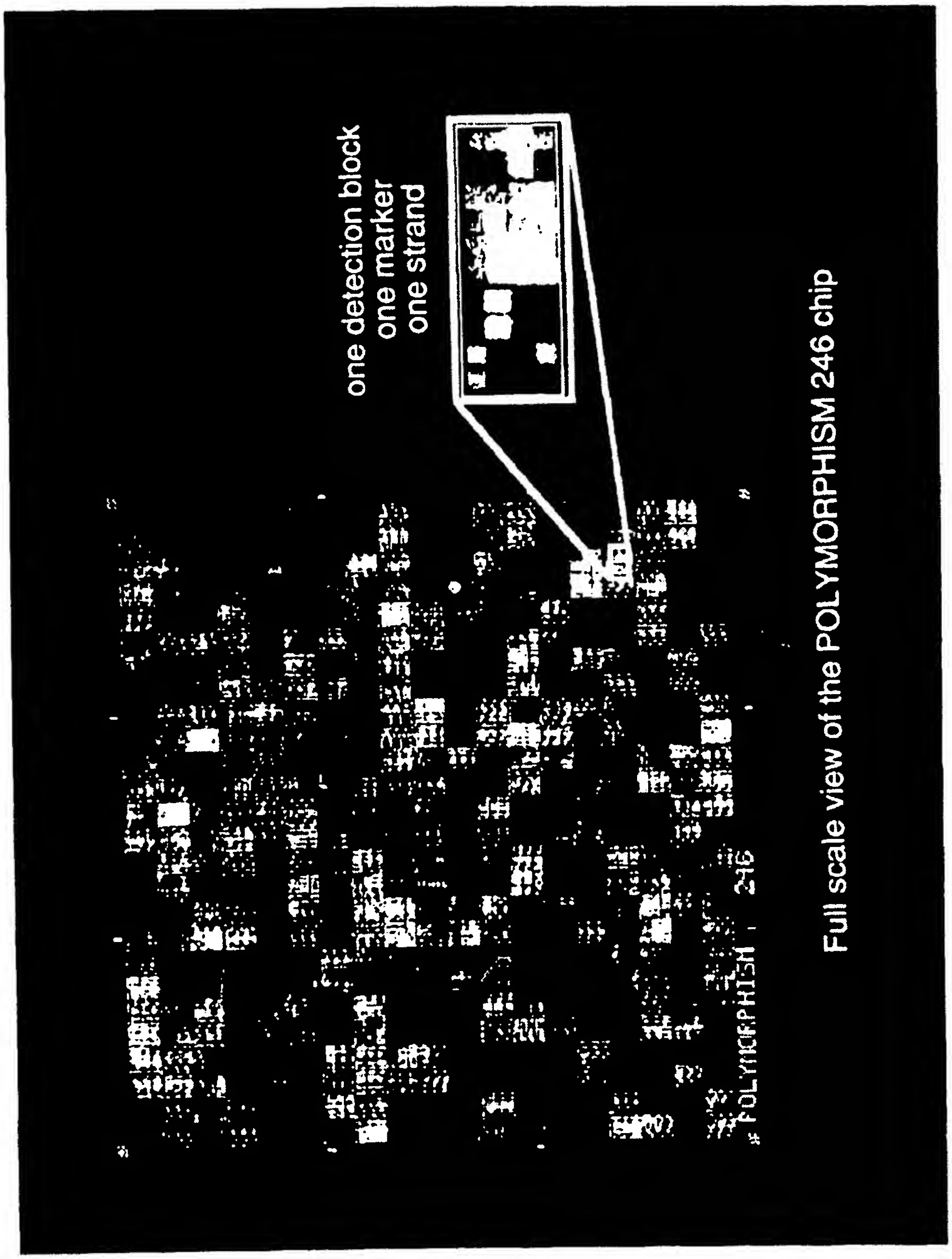

one detection block
one marker
one strand
Full scale view of the POLYMORPHISM 246 chip
POLYMORPHISM 246

FIG. 10B